Running title: Gesture delay and coda /r/ weakening.

Title: The role of gesture delay in coda /r/ weakening: an articulatory, auditory and acoustic study.

Eleanor Lawson[1], Jane Stuart-Smith[2], James M. Scobbie[1]

[1] *CASL Research Centre, Speech and Hearing Sciences, Queen Margaret University, Edinburgh, EH21 6UU, U.K.*

[2] *GULP lab, English Language, 12, University Gardens, University of Glasgow, Glasgow, G12 8QQ, U.K.*

Abstract

The cross-linguistic tendency of coda consonants to weaken, vocalize, or be deleted is shown to have a phonetic basis, resulting from gesture reduction, or variation in gesture timing. This study investigates the effects of the timing of the anterior tongue gesture for coda /r/ on acoustics and perceived strength of rhoticity, making use of two sociolects of Central Scotland (working-class and middle-class) where coda /r/ is weakening and strengthening respectively. Previous articulatory analysis revealed a strong tendency for these sociolects to use different coda /r/ tongue configurations – working-class and middle-class speakers tend to use tip/front raised and bunched variants respectively; however, this finding does not explain working-class /r/ weakening. A correlational analysis in the current study showed a robust relationship between anterior lingual gesture timing, F3, and percept of rhoticity. A linear mixed effects regression analysis showed that both speaker social class and linguistic factors (word structure and the checked/unchecked status of the prerhotic vowel) had significant effects on tongue gesture timing and formant values. This study provides further evidence that gesture delay can be a phonetic mechanism for coda rhotic weakening and apparent loss, but social class emerges as the dominant factor driving lingual gesture timing variation.

Keywords: articulatory phonetics, ultrasound, acoustics, sociophonetics, rhotics

I. INTRODUCTION

A. Phonetic basis for coda consonant weakening

The notion that there is a phonetic basis for the cross-linguistic tendency of coda consonants to lenite, vocalize or be deleted has, for a long time, been of interest to phoneticians and phonologists. Articulatory studies of speech have identified two key mechanisms that can underlie coda consonant weakening: articulatory reduction - a decrease in the magnitude of articulatory gestures, and variation in gesture timing - change in the timing of articulatory gestures relative to other speech events.

Browman and Goldstein's (1995) x-ray microbeam study of American English consonants observed gestural reduction by tracking the position of pellets attached to the tongue blade and showed a reduction in height when /l/, /t/ and /n/ were in coda position, in comparison to onset position. They also observed a reduction in lip constriction for /p/ when it was in coda rather than onset position. They suggested that this phenomenon might be caused by a general reduction in speaking effort over the time course of the syllable (Browman and Goldstein, 1995).

A large body of articulatory research, much of it focused on /l/, uncovered another mechanism that potentially contributes to coda consonant weakening, namely variation in the synchronicity of the primary and secondary lingual gestures (Sproat and Fujimura, 1993, Recasens and Farnetani, 1994, Krakow, 1989). Sproat and Fujimura's study of American English /l/, using x-ray microbeam, found that coda /l/ darkness correlated strongly with the acoustically measured duration of the rime containing the coda /l/. It was found that the stronger the following phonological boundary, the longer the preceding syllable rime and the darker the /l/, quantified in terms of F2-F1. Sproat and Fujimura initially hypothesised shorter rimes led to articulatory undershoot, whereby the tongue dorsum failed to retract as much as it would be able to in a longer rime; however, it was also found that longer rimes were associated with a greater durational difference between the primary apical gesture and the secondary dorsal gesture for /l/. In other words, both gestural undershoot and increased dissociation of gestures contributed to a darker, more vocalised, quality. The boundary that conditioned both the greatest degree of tongue dorsum retraction and the greatest degree of temporal separation between apical and dorsal gestures was the ‘major intonation boundary’, i.e. /l/ in pre-pausal position.

The possibility that variation in gesture synchrony could account for more than changes in phonetic quality, i.e. could be a mechanism for diachronic segment deletion has been suggested by Recasens and Farnetani (1994), who carried out an EPG study of contact patterns for phrase-initial and phrase-final /l/, produced by single speakers of American English, Catalan and Italian. They noted that the alveolar gesture of phrase-final dark /l/ in Catalan and American English not only occurred later than the dorsal gesture, but was found to occur partially or completely after the offset of voicing, leading to apparent consonant deletion at the auditory and acoustic levels, but not at the articulatory level, i.e. an apical articulatory gesture was present that was “devoid of acoustic consequences” (Recasens and Farnetani, 1994: 204). Although Recasens and Farnetani concluded that loss of /l/ must be production-based rather than perception-based, Browman and Goldstein have suggested that if an anterior lingual gesture occurs in utterance-final silence, then deletion of that gesture might then become a listener-based sound change, (Browman and Goldstein, 1995: 26), see also (Ohala, 1981), i.e. listeners would reinterpret the auditorily covert gesture as deletion and might fail to produce the gesture at all in their own speech. Lawson, Scobbie and Stuart-Smith’s studies of articulatory adaptation during mimicry of audio stimuli show that misinterpretation of weakly audible rhotics with covert lingual gestures can occur, but not as often as might be expected. Top-down information, such as lexical access, probably reduce instances of misinterpretation (Lawson et al., 2011, Lawson et al., 2014b), see §IV.A1 for further discussion of misinterpretation of covert gestures.

The present study uses ultrasound tongue imaging to investigate the role of lingual gesture timing in the audible weakening of coda /r/, which has received less attention than /l/ to date. Most studies look at liquid consonant gesture timing in highly-constrained sets of utterances (Sproat and Fujimura, 1993, Scobbie and Pouplier, 2010, Turton, 2014), often involving the flanking of the liquid under study with front high vowels in order to be able to reliably distinguish between tongue gesture movements associated with the consonant and those associated with the vowel. Posterior articulatory tongue gestures for /r/ (i.e. tongue root retraction gestures) tend to have merged with those of preceding non-high vowels and therefore cannot always be reliably identified. As rhoticity strength shows such strong social stratification in the communities under study and because we did not want participants to be aware that the study concerned /r/, we included words with a wide range of (i.e. both high and non-high) prerhotic vowels. This meant that rather than looking at the timing of the anterior (tip or dorsum raising) and posterior (root retraction) gestures that make up coda /r/, we could only reliably identify the anterior gestures. We opted to quantify the timing of the anterior lingual gesture relative to the offset of voicing, or onset of a following labial consonant; two events that have the potential to audibly mask the anterior /r/ gesture.

B. Rhoticity in Central Scotland

In this study we make use of socially-stratified variation in coda /r/ production that is evident in different sociolects in Central Scotland, in order to study the role of tongue gesture timing variation in /r/ weakening, and how it relates to variation in the acoustic characteristics of /r/ and variation in perceived strength of rhoticity.

For several decades, sociolinguistic researchers have noted weakening of coda /r/ in the vernacular English of Central Scotland (Romaine, 1978, Speitel and Johnston, 1983, Macafee, 1983, Stuart-Smith, 2003, Stuart-Smith, 2007a, Jauriberry et al., 2012). These mainly auditory-acoustic studies have shown that strength of rhoticity is socially stratified, with middle-class speakers preserving, and even strengthening rhoticity, while working-class speakers are often weakly rhotic. By weakly rhotic, we mean that minimal pairs such as *bud/bird* /bʌd/bʌrd/ and *cod/cord* /kɔd/kɔrd/ can, for the most part, still be differentiated by local speakers, but are not easily differentiated by those unfamiliar with working-class Central Scottish English (see Lennon, 2013), /r/ is so weakly produced that *cod* and *cord* can sound equally /r/-less to those not familiar with the Central Scottish accent. Acoustic and auditory analyses identify some prerhotic vowel modifications that are associated with coda /r/ weakening, such as pharyngealisation of the prerhotic vowel [bʌˤd] "bird", or lengthening of the prerhotic vowel. Stuart-Smith's auditory-acoustic study of working-class Glaswegian postvocalic /r/ found that if /r/-ful words were derhoticised, formant values and trajectories for /r/-ful and /r/-less minimal pairs were very similar, but that an acoustic correlate of pharyngealised or uvularised derhotic variants appeared to be F3 raising. Stuart-Smith also found instances of /r/-ful words where there were no changes in formant frequencies or amplitudes throughout the vocalic portion of the vowel, e.g. "heart" produced as [ha̠ʔ]. Finally, it was found that the vocalic portion of derhoticised /r/-ful words was significantly longer than that of /r/-less words (Stuart-Smith 2007). Stuart-Smith (2007) concluded that "acoustic analysis shows few straightforward links with auditory findings" (Stuart-Smith 2007: 1452) and that "explaining (the derhoticisation) process will need recourse to articulation." (ibid.). In an acoustic analysis of /r/-ful and /r/-less minimal pairs in Glaswegian English, Lennon et al. (2015) found that working-class /-id/ /-ird/ minimal pairs were easy to differentiate due to the presence of a prerhotic offglide:

[bid] “bead” versus [biʌd] “beard”. However, formant tracks throughout the vocalic portion of the word for /-ʌd/ /-ʌrd/ minimal pairs showed a great deal of similarity, with significant variation found only in the duration of the vocalic portion of the word (/r/-ful tokens were longer) and variation in F2 (F2 was found to be lower throughout the vocalic portion for /r/-ful tokens). It would seem therefore that articulatory analysis is required in order to fully understand Central Scottish derhoticisation.

To date, articulatory analysis of coda /r/ in Central Scottish speech communities has already uncovered articulatory variation of which sociolinguistic and dialectological researchers were unaware. There is a strong tendency for coda /r/ tongue shape to be socially-stratified, with working-class speakers tending to produce coda /r/ with tongue-tip/front raised variants of /r/, while middle-class speakers tend to produce /r/ with tongue bunching (Lawson et al., 2011, Lawson et al., 2014a). Bunched /r/ variants have been identified and studied along with retroflex /r/ for decades in American English, see (Delattre, Pierre & Freeman, Donald C., 1968, Lindau, 1985, Mielke et al., 2010, Zhou et al., 2008). In studies of American English, these /r/ tongue shape variants appear to be used idiosyncratically, or for coarticulatory reasons, e.g. bunched variants occurring before high front vowels. In Central Scottish English, socially-stratified variation between i) tip/front raised and ii) bunched coda /r/ variants has been identified in four separate adolescent ultrasound word list speech corpora, collected between 2007 and 2014 (Lawson et al., 2014b, Lawson et al., 2014a, Lawson et al., 2011), including the dataset used in the current study (Lawson et al., 2014a). Middle-class speakers in the current study used mainly bunched /r/ variants, while working-class speakers mainly used tip-up and front-up variants. Fig. 1 below illustrates tongue shape variation in the current dataset, with an average 94% of coda /r/ tongue shapes produced by middle-class speakers classified as bunched, while 94% of coda /r/ tongue shapes produced working-class speakers were classified as “tip up” or “front up”.

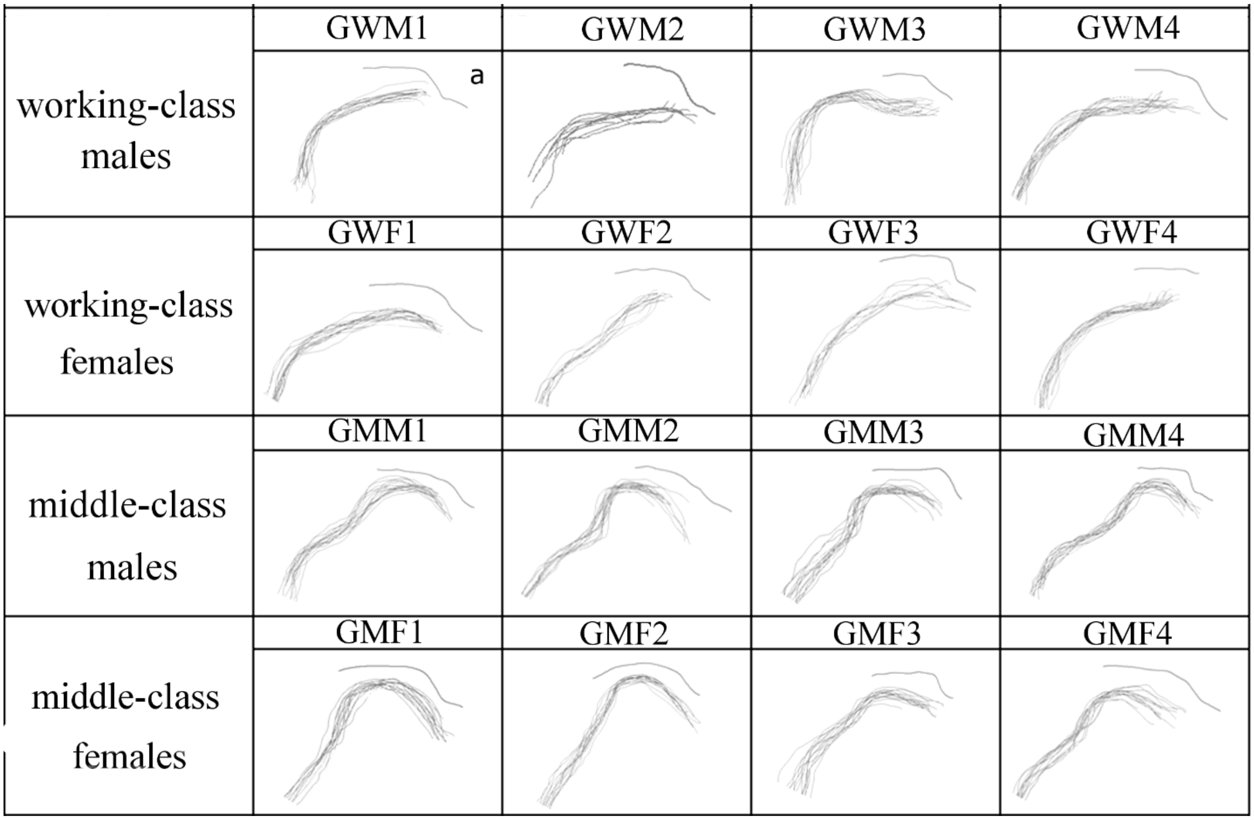

FIG. 1: Bundles of CVr-word midsagittal tongue surface splines, extracted from the tongue surface at the point of maximum constriction for /r/, organized by social class and sex. The uppermost line in each cell represents the midsagittal alveolar ridge and hard palate, with the alveolar ridge to the right of the cell. (a) GWM1 produced only tapped and trilled /r/ variants and was excluded from the study. Figure adapted from Fig. 12-8 (Lawson et al., 2014a).

The social stratification of tongue configurational variation for coda /r/ might explain some of the audible differences in rhotic strength between working-class and middle-class coda /r/, in that bunched /r/ variants might produce a particularly strong audible impression of rhoticity. Delattre and Freeman, for example, found that a bunched tongue shape type 4, ‘dorsal bunched with dip’ (Delattre and Freeman 1968: 41) produced the “strongest auditory impression” (ibid.: 64) of rhoticity. Nevertheless, tongue configuration variation does not explain the extreme degree of weakening found in working-class Scottish coda /r/. A tip/front raised /r/ tongue shape is *not* associated with weak rhoticity per se, and equally rhotic-sounding tip/front-raised and bunched variants of /r/ are found in other varieties of English (see Delattre, Pierre & Freeman, Donald C., 1968, Zhang et al., 2003, Twist et al., 2007). In other UTI-acoustic datasets collected in Central Scotland, a delay in the tongue tip/front raising gesture has been observed and reported in working-class speech (Lawson et al., 2011, Lawson et al., 2014a), but never studied systematically, or quantified until now. The dataset used in this study therefore allows us to determine whether lingual gesture timing variation underlies the natural weakening of coda /r/, by analyzing covariation between (a) the timing of underlying tongue gestures, (b) percept of rhoticity and (c) formant patterns in two speaker groups who weaken and strengthen coda /r/ respectively. It also allows us to determine whether speaker social class is a predictor of rhotic gesture timing.

We hypothesise that underlying the weakly-rhotic, working-class coda /r/ is a delay in the timing of the tip/front raising gesture, to the extent that it is partly or fully masked by silence after voicing offset, or by the frication or closed phase of a following labial consonant. We also hypothesise that variation in lingual gesture timing will correlate with variation in the auditory percept of rhoticity and we will identify the acoustic correlates of tokens with early and delayed anterior /r/ gestures.

II. METHOD

A. Participants

The Western Central Belt audio-ultrasound tongue imaging corpus (henceforth WCB12) was collected in 2012 in the city of Glasgow, Scotland. We recorded young adolescents (12-13 year olds), as younger speakers tend to produce the best-quality ultrasound images due to their smaller head size, i.e. there is a shorter distance from the ultrasound probe surface to the tongue surface. Recruiting informants from schools also helped socially stratify our corpus, as schools with affluent vs. deprived catchment areas were approached to participate.

Sixteen Glaswegian speakers aged 12-13 were recorded for this study. Half were male and half were female, from schools that were geographically close to one another (within two miles). Demographic information, presented in Table I, indicates pupils' different social backgrounds and potential future socio-economic trajectories.

TABLE I: Key (2011/12) social demographic information pertaining to the two schools involved in the study. (a) data for 2010/11.

	Working-class school	Middle-class school
16 yr old students obtaining 5 or more awards at highest grade	28%	79%
Students going on to Higher education	22% [a]	77% [a]
Students registered for free school meals[1]	30%	3%

B. Recording scenario and equipment

Informants were recorded with audio and ultrasound tongue imaging (UTI) in an IAC sound-attenuated recording booth at the University of Glasgow. All noisemaking equipment such as the ultrasound machine and PC were located outside of the recording booth. Participants were fitted with a stabilising headset with the ultrasound probe held in place underneath the chin, to eliminate roll and yaw, and minimise pitch movement of the ultrasound probe in relation to the head (Scobbie et al., 2008).

Single word prompts were presented orthographically to participants one at a time on a monitor. Audio recordings were made using a Beyer-Dynamic Opus 55 headworn microphone. Audio recordings were sampled at 22kHz. A Mindray DP2200 ultrasound machine, set to NTSC video format, created UTI video at a target rate of 29.97fps. The frame rate of the UTI video was doubled to 59.94fps by deinterlacing each video frame post hoc. See §II.D for details of audio-video synchronisation.

C. Word list

Our word list had 30 stressed, monosyllabic items containing coda /r/, 25 with a CVr structure, e.g. *bear* and 5 with a CVrC structure, e.g. *herb*, see Table II. There were also 14 /r/-ful nonsense words (not analysed in this study) and 98 distractors. 9 words in the word list contained one of the set of Scottish checked vowels /ɪ, ʌ, ɛ/, which are of particular interest in relation to gesture timing in /r/. The term “checked” refers to the phonotactic specification that vowels do not occur in stressed open word-final syllables, i.e. they must always be followed by a consonant in stressed syllables. Checked vowels tend to be phonetically more lax than unchecked vowels and are also shorter than unchecked vowels in most varieties of English, though not in Scottish English where vowel length is not phonemic (Scobbie et al., 1999). The short, lax phonetic quality of the checked vowels seems to have permitted following /r/ to exert a strong coarticulatory force over them historically. In American English, we see historical coalescence of these vowels with /r/ and merger of the three vowels to [ɚ] (Wells, 1982a: §6.1.5). Historical changes in Anglo-English relating to the checked vowels are not recoverable, but in present-day Anglo-English, these vowels are merged to [ɜː], known as the NURSE merger (Wells, 1982b: §3.1.8), which could have entailed an initial merger/coalescence to [ɚ], as in American English, followed by loss of the rhotic colouring of the vowel, or a merger to a central vowel before /r/, followed by loss of the rhotic segment. In Scottish English, merger/coalescence of /ɪ, ʌ, ɛ/ before /r/ to [ɚ] is a longstanding feature of middle-class, but not working-class speech, and previously thought to be an adaptation towards Anglo-English phonology (see Aitken, 1979). However, and ultrasound-based study by Lawson et al. (2013) has shown that coalescence of the checked vowels and /r/ occurs due to the strong coarticulatory force exerted by bunched /r/. Lawson et al. (2013) suggested that the timing of the maximum of the rhotic gesture occurs early in /ɪr, ɛr, ʌr/ syllables where bunched /r/ is used, and not when tip-up/front-up /r/ is used, but this vowel-dependent rhotic timing variation has not been quantified to date. The inclusion of checked and unchecked prerhotic vowel variants in this study, and the inclusion of the fixed factor PRERHOTIC VOWEL with levels checked and unchecked in our statistical analysis, will allow us to quantify the impact of the checked status of the vowel on the gestural timing of coda /r/.

TABLE II: Word list items, arranged according to prerhotic vowel.

Checked prerhotic vowel /ɪr, ʌr, ɛr/		**Unchecked prerhotic vowels**					
		/ir/	**/er/**	**/ar/**	**/ɔr/**	**/or/**	**/ʉr/**
fir	err	ear	air	bar	for	bore	boor
firm	her	beer	bear	far	form	more	moor
burp	herb	fear	hair	par	or	oar	poor
fur	perm	peer	pair			pore	
purr							

There is some historical evidence that coda /r/ was lost earliest in preconsonantal position in Anglo-English (Dobson, 1957: §401). Structure of the word, i.e. CVr or CVrC structure, as in “fir” and “firm”, was therefore included in the statistical analysis of the dataset, as the fixed factor STRUCTURE, in order to identify whether timing of the anterior rhotic gesture is affected by a word’s structure, or whether acoustic and auditory measures are affected by a word’s structure. For example, it is possible that delayed gestures are less audible in CVrC words than in CVr words followed by silence, as the final consonant in CVrC words provides more absolute masking of the /r/ gesture than voicing offset does.

We avoided lingual consonants in prompts, to reduce potential coarticulatory effects on /r/. Working-class informants occasionally produced disyllabic versions of the words *form, firm,* and *perm* with a tapped /r/ variant [fɔɾʌm], [fɪɾʌm], [pʰɛɾʌm]. These tokens were excluded.

D. Ultrasound-audio resynchronization

An essential preliminary phase was the resynchronization of the audio and ultrasound video channels and calculation of internal processing lag for the video-based ultrasound machine. The audio-ultrasound recording system, like many others, involves separate audio and video channels, received and processed by a laptop computer. We therefore needed a post hoc means of synchronizing audio and video and re-establishing the video frame rate. Both audio and video channels passed through a SynchBrightUp unit (Articulate Instruments Ltd.) which acted as a clapperboard system, superimposing a flash on the video signal and a tone on the audio signal at the beginning of each new recording and encoding information about video frame rate. These signals were later used by Articulate Assistant Advanced (AAA) ultrasound recording and analysis software (Wrench, 2012) to re-establish the UTI video frame rate and to resynchronise audio and video on each recording by aligning flash and tone. Additionally, video-output ultrasound machines have a variable internal processing delay of several milliseconds, while the data collected at the probe is turned into a video frame.

The mean internal video processing delay of an ultrasound scanner can be estimated using a 'tap test': the microphone capsule was tapped onto the ultrasound probe, and the mean delay from the audio to the visual record of this event was calculated from 100 taps. The DP2200 was found to have an average image processing delay of 20ms (standard deviation = 14ms) - a little under one deinterlaced video frame. A -20ms lag was introduced to the video signal to account for this delay. The variability of the processing delay means that there is slight random variation in the amount of time it takes the DP2200 to create and output each video frame. Such inconsistencies in synchronization of video and audio act as randomly-distributed noise in the data.

E. Tongue gesture timing annotation

As already mentioned, the present study quantifies the durational difference between the maximum of the anterior rhotic lingual gesture in coda /r/ (tip/front raising or dorsum raising, depending on the type of /r/ involved), and either voicing offset or following consonant onset. In order to measure this durational difference and normalize the measure, four main temporal events in each coda /r/ token were annotated:

(1) *rmax* - the visually-determined location of the maximum of the anterior lingual /r/ gesture.

(2) *V-onset* - the acoustically-determined location of the onset of the vowel in CVr and CVrC words.

(3) *voice-offset* - the acoustically-determined location of the offset of voicing in CVr words.

(4) *C-onset* - the acoustically-determined location of the onset of the closing labial consonant in CVrC words.

These key temporal measurements reveal the relationship between the timing of *rmax* and articulatory events that could render the /r/ gesture partly or completely inaudible (Fig. 2).

V-onset, *voice-offset* and *C-onset* values were annotated by the first author using Praat (Boersma, Paul & Weenink, David, 2013), using waveform, spectrogram and Praat tools such as the pitch and intensity trackers as a guide.

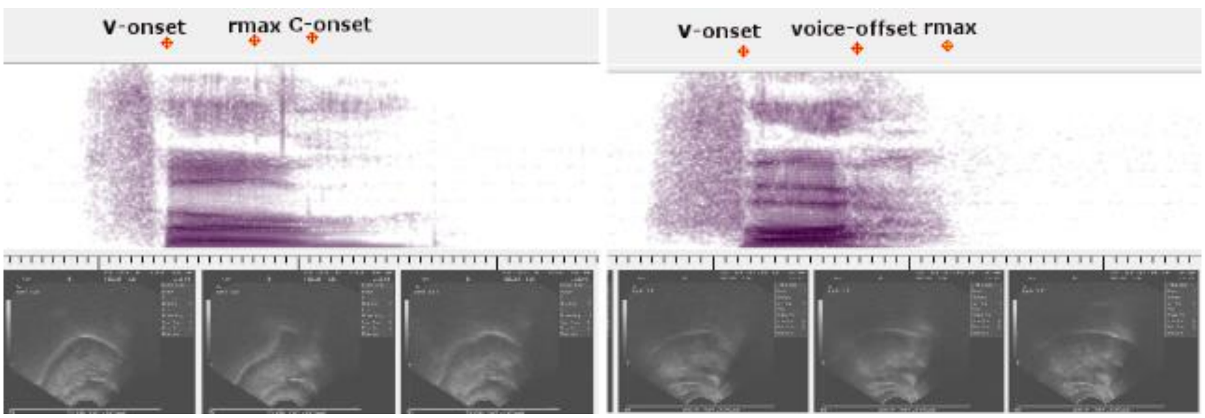

FIG. 2 (Left) MC speaker (GMM2) "firm" [fɚm] showing V-onset, rmax and C-onset annotations with spectrogram. (Right) WC speaker (GWF4) "fir" [fʌˤ] showing V-onset, rmax and voice-offset annotations with spectrogram. Tickmarks below show the alignment of the UTI keyframes with the acoustic signal, below are sample UTI images showing the progression of the articulations through the recording.

Raw lag duration between *rmax* and an articulatory event that could render the /r/ gesture inaudible was calculated as follows:

CVr words *ear, bar, pore* etc. (lag = *rmax* – *voice-offset*)

CVrC words *herb*, *firm*, *form* etc. (lag = *rmax* – *C-onset*)

Time-normalisation of these raw measurements was also carried out to take account of potentially different speech rates. Normalisation involved dividing the *raw lag* durational measure by the *V-onset* to *voice-offset*/*C-onset* duration, which means that *normalised lag* is *raw lag* expressed as a proportion of the vowel + /r/ section of the syllable rime. Where *rmax* occurs before voice-offset or C-onset, as in Fig. 2, left, *raw lag* and *normalised lag* are negative. Where *rmax* occurs after voice-offset or C-onset, as in Fig. 2, right, *raw lag* and *normalised lag* are positive.

F. Auditory analysis

In order to confirm the hypothesis that delayed anterior /r/ gestures are associated with weakened audible rhoticity, all tokens of /r/ were rated on a rhoticity index, using a Praat *multiple forced choice* (MFC) experiment interface. The interface presented randomised anonymous tokens to three classifiers (the authors), two speakers of Standard Scottish English (SSE) from the Central Scotland, and one speaker of Standard Southern British English (SSBE) with twenty years' experience studying Central Scottish speech. Each token was rated on a rhoticity index – 5-point ordinal scale, arranged from /r/-less to strongly /r/-ful. We used an /r/-index of ordered categories ranging from least to most /r/-ful to obtain greater consistency between raters and hence gain a more meaningful capture of their ratings. The r-index itself is based on gradient r-indexes in Scottish rhoticity literature, e.g. (Stuart-Smith, 2007b; Jauriberry et al., 2012). Category labels registered in the MFC as numerical codes ranging from 1 (no /r/) to 5 (schwar). Despite tapped and trilled variants being stereotypically associated with Scottish speech, only one speaker, GWM1, produced them, but never in his spontaneous speech. He was also the only speaker who identified /r/ as the focus of the study, therefore this speaker was removed from the study.

In the rhoticity index, henceforth *r-index, no /r/* referred to no auditory percept of /r/, i.e. the word sounded as if it ended in a (non-rhoticised) vowel e.g. [fʌ] "fur", [fiɐ] "fear"; *derhoticised* referred to variants where there was an audible hint of /r/, or some other feature that could be associated with rhoticity in Glaswegian English, such as pharyngealisation or velarisation of the prerhotic vowel, but no clear rhotic segment e.g. [fʌˤ], [fiʌˤ] [2]; *alveolar* referred to a (post)alveolar approximant with a less strong rhotic quality e.g. [fʌɹ], [fiəɹ] than the *retroflex* approximant, which referred to a variant that sounded like a strongly rhotic approximant e.g. [fʌɻ], [fiəɻ] and, finally, *schwar* was a central rhotic vowel [ɚ] in place of a Vr sequence e.g. [fɚː], [fiɚ]. From a phonological point of view, [ɚ] might be considered by some to be a vocalisation of /r/ and therefore weaker than a retroflex approximant; however, in this study [ɚ] was considered by the raters to have very strong audible rhoticity. This phonetic variant is particularly associated with an underlying bunched articulation of /r/, occurring after the checked vowels /ɪ, ɛ, ʌ/ in Scottish English (see Lawson et al., 2013). As [ɚ] represents a coalescence of the vowel and /r/, it is highly likely to involve an early anterior lingual /r/ gesture, close to the syllable nucleus.

(C)Vr(C) word tokens were classified by each auditory rater over several sessions and a mean rhoticity index value was calculated for each token. Inter-rater reliability was gauged using Krippendorff's alpha (Hayes and Krippendorff, 2007) using the *irr* package (Gamer et al., 2012) in R (Core Team, 2013), showing a moderate level of inter-rater reliability between the three raters α=0.754. As the mean *r-index* scores are ranked ordinal data, rather than interval data, we did not include *r-index* in the mixed effects modelling; however, it is included in a nonparametric correlational analysis, §III.A. We carried out nonparametric Kruskall-Wallis tests on *r-index* with CLASS, SEX, PRECEDING VOWEL and STRUCTURE as fixed factors, see §III.D

G. Acoustic analysis

Wherever the articulatorily-determined *rmax* occurred before the offset of voicing, or onset of a following consonant, formants one to five were measured by hand using Praat at the same temporal point of the articulatory *rmax* annotation. Measurement of the first five formants involved inspection of the spectrogram alongside narrowband FFT and LPC spectra. Close inspection of spectra was required, as, for many strongly /r/-ful variants, F2 and F3 were not easy to differentiate using only a spectrographic representation, due to their proximity. Where *rmax* occurred after *voice-offset/C-onset*, formants one to five were measured just before *voice-offset/C-onset.* The first five formants were measured, rather than just the traditional first three formants for rhotics. Although F1-F3 measures are traditionally considered sufficient to characterise rhotic variability, an MRI-based study by (Zhou et al., 2008), albeit one based on productions of /r/ by only two American speakers, suggested that retroflex and bunched variants may be acoustically distinct and auditorily discriminable due to variation in the higher formants, specifically F4 and F5, with bunched /r/ being characterised by a greater acoustic distance between F3 and F4, and a lesser acoustic distance between F4 and F5, than retroflex /r/. Given the socially-stratified bunched-retroflex variation in our dataset (Lawson et al., 2014a), by measuring F4 and F5 in addition to F1-F3 we aimed to take account of the potential impact of tongue shape variation on the acoustic signal and to determine whether tongue shape was also having an impact on perceived strength of rhoticity. In addition, we hoped to contribute to the understanding of the acoustics of bunched-retroflex variation in /r/. However, we encountered difficulties in measuring F5, particularly at voicing offset and no measurement could be taken for around 10% of tokens, mainly tokens from working-class speakers. We were not completely confident of the accuracy of F5 measures and initial linear effects modelling showed a lack of significant findings for F5 for both fixed and random factors, therefore we decided to analyse F1-F4 only.

Whilst differences in vocal tract size between speakers are likely to result in small differences in formant values for /r/ for this early adolescent speaker group, just as they would for vowels (e.g. Adank et al., 2004), there is currently no accepted method of acoustic normalisation for rhotics, particularly the higher formants. In our mixed effects regression analysis (see §II.H) we included the fixed effect of SEX with two levels: male and female, and we also included SPEAKER as a random factor. Arguably, given that the speakers were aged 12-13, the inclusion of the random factor SPEAKER may be more important than the fixed factor SEX, as in this early stage of adolescence, sex-based vocal tract differences are less predictable. Anecdotally, the authors often misidentified males as females and vice versa in the study when listening to audio recordings.

H. Statistical analysis

In order to examine the relationship between the articulatory, auditory and acoustic variables in the study, and to test the hypothesis that gesture delay is responsible for coda /r/ weakening, we first carried out a Spearman's correlational analysis of all the dependent measures with Bonferroni corrections to take account of the fact that a series of multiple tests were run simultaneously.

Thereafter, given that coda /r/ weakening and strengthening are particularly associated with working-class and middle-class speech respectively, we used mixed effects modelling to determine whether social class was a significant predictor of variation across the articulatory and acoustic measures taken. In these models, we also took account of other features of our dataset that might potentially have a significant effect on the dependent measures. Our six dependent variables were: *raw lag*, *normalised lag* and *F1 to F4*. The following four fixed factors were included: (1) PRECEDING VOWEL (checked /ɪ/, /ɛ/, /ʌ/, or unchecked /i/, /ʉ/, /a/, /o/, /ɔ/, /e/), (2) STRUCTURE of the word (CVr e.g. *moor* or (C)VrC e.g. *form*), (3) CLASS (working-class (WC) or middle-class (MC)) and (4) SEX (male or female). We also tested for interactions between CLASS and SEX in order to identify whether males and females were behaving differently within their class groups, as has been suggested (Jauriberry et al., 2012, Lawson et al., 2014a), and between CLASS and PRECEDING VOWEL, because we expected vowel-to-/r/ coalescence where /r/ follows checked vowels in the middle-class group, which we would expect to have an impact on all of the dependent variables. A variance-inflation-factors test was carried out in order to check for collinearity among fixed factors, where any factor obtaining a value > 2 would be removed. All fixed factors were kept.

Mixed effects modelling takes account of aspects of the experimental design that involve sampling within a population, allowing the inclusion of random factors that are not generalizable to the wider population in the way that fixed factors such as CLASS, SEX etc. would be. SPEAKER and PROMPT were included as random factors to prevent extremes of variation in the behavior of particular speakers or extremes of variation in the production of particular prompts having an undue effect on statistical results, see (Drager and Hay, 2012, Hay, 2011). The inclusion of SPEAKER as a random factor was particularly important in the construction of models for the formant data, as it helped to remove formant variation that might be attributable to variation in vocal-tract length. SPEAKER was found to be significant for all factors. PROMPT was significant for *F1 – F3* only. We used the lme4 package in R (v3.1.2) followed by the step() function to find the models that best fit the data. The auditory measure, the mean *r-index* scores, constitute ordinal rather than interval data. We carried out nonparametric Kruskall-Wallis tests on *r-index* with CLASS, SEX, PRECEDING VOWEL and STRUCTURE as fixed factors.

We will only report on effects and correlations that were found to be significant.

III. RESULTS

A. Correlational analysis

We analysed the relationship between the articulatory, auditory and acoustic measures taken from the productions of /r/ from this socially stratified speaker sample. A series of Spearman's correlation tests with Bonferroni corrections were carried out on the dependent variables *raw lag*, *normalised lag*, *r-index* and *F1-F4*. Most dependent variables were found to correlate with one another, see Table III.

TABLE. III: Correlation matrix for working-class speakers, showing Spearman's *r*. *** = significant to the $p<0.001$ level, * = significant to the $p<0.05$ level. Bold font indications correlations where $-0.5 \geq r_S \geq 0.5$.

	raw	norm	rindex	F1	F2	F3	F4
raw		**0.98*** **	**-0.69*** **	0.21	**-0.50*** **	**0.57*** **	-0.30***
norm			**-0.69*** **	0.18	-0.48***	**0.57*** **	-0.28***
rindex				-0.25***	0.44***	**-0.65*** **	0.30***
F1					-0.08	0.25***	-0.14*
F2						-0.06	0.09
F3							-0.31***
F4							

We inspected and interpreted only the strong correlations, which we define here as those which were not only significant, but also showed $-0.5 \geq r_S \geq 0.5$. These strong correlations are displayed in the scatterplots in Figure 3.

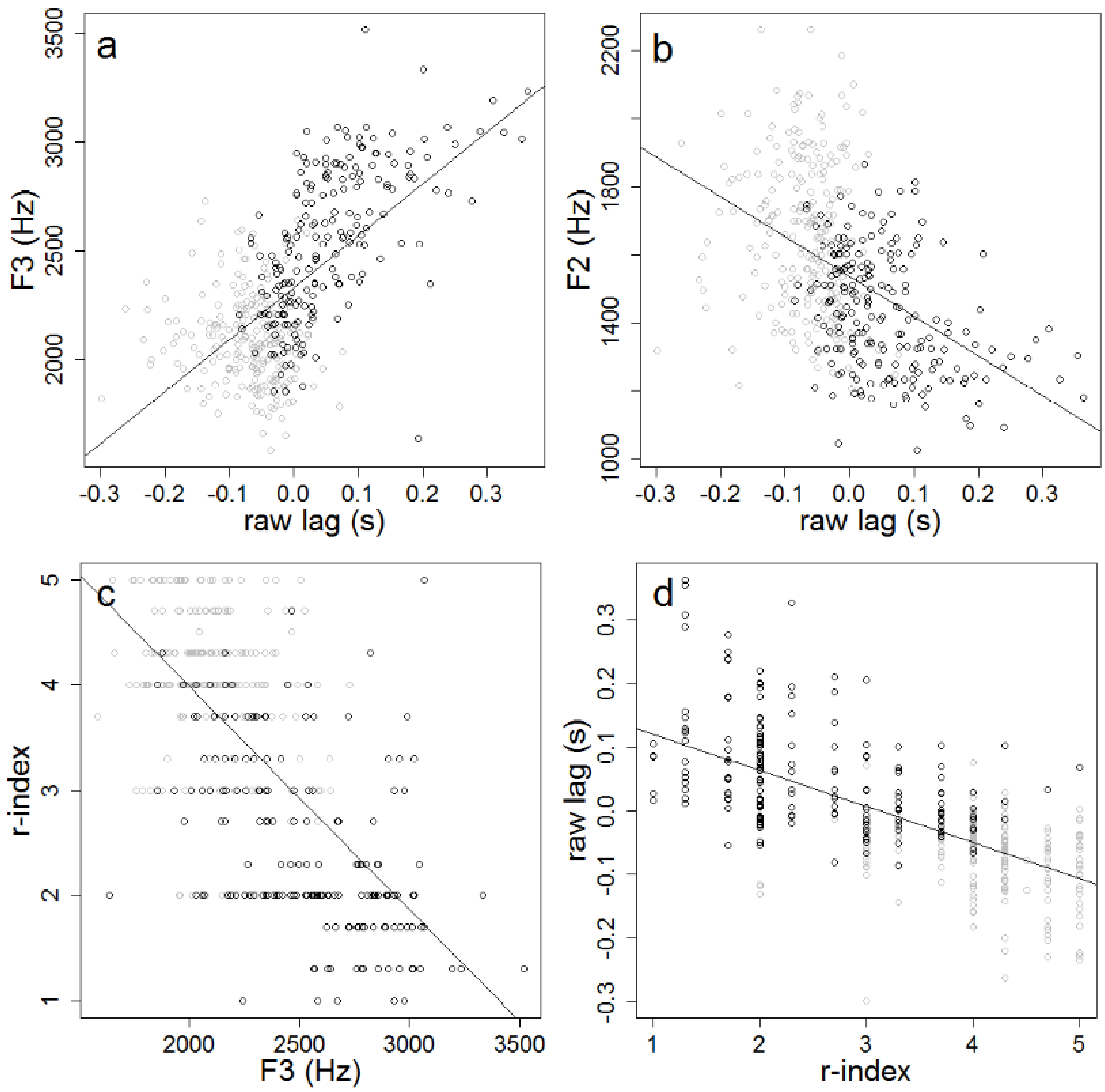

FIG. 3: Scatterplots of strongest correlations in working-class dataset: (a) raw lag and F3 (b) raw lag and F2 (c) F3 and r-index (d) r-index and raw lag. Working-class data points are black, while middle-class data points are grey.

The two articulatory measures, *raw* and *normalised lag* were, unsurprisingly, the most closely correlated $r_S = 0.98$, $p<0.001$. After this, the correlation matrix (Table III) and plots (Fig. 3) reveal a correlational triangle between articulatory and auditory variables and the acoustic variable most strongly associated with rhoticity, *F3*, and to a lesser extent, *F2*. There were four key findings from the correlational analysis: (1) There was a strong negative correlation between *r-index* and *lag* ($r_S = -0.69$, $p<0.001$), see Fig. 3d; the greater the *lag*, the less /r/-ful sounding the word. (2) There was a strong negative correlation between *F3* and *r-index* ($r_S = -0.65$, $p<0.001$), see Fig. 3c; the higher the *F3*, the less /r/-ful sounding the word. (3) There was a strong positive correlation between *lag* and *F3* ($r_S = 0.57$, $p<0.001$), see Fig. 3a; the greater the *lag*, the higher the *F3*. (4) There was a strong negative correlation between *F2* and *raw lag* ($r_S = -0.50$, $p<0.001$), see Fig. 3b; the greater the *raw lag*, the lower the *F2*. In summary, our hypothesis was correct; the greater the *lag*, the higher *F3* and the less /r/-ful sounding the word.

The correlation plots (Fig. 3) also illustrate social stratification of the dependent variables *raw lag*, *r-index*, *F2* and *F3*, as the working-class speakers' tokens, black circles, are shown to have shorter lags, lower *r-index* scores and lower *F2*s and higher *F3*s than the middle-class speakers' tokens, grey circles. The following sections confirm the impact of social class on these measures, also in interaction with linguistic factors.

Other significant, but weaker, correlations were found where $-0.5 < r_S > 0.50$. *F2* was positively correlated with *r-index* ($r_S = 0.44$, $p<0.001$); the higher *F2*, the more /r/-ful sounding the word. *F3* was negatively correlated with *F4* ($r_S = -0.31$, $p<0.001$); the higher *F3*, the higher *F4*. *Raw lag* was negatively correlated with *F4* ($r_S = -0.30$, $p<0.001$); the greater the *lag*, the lower *F4*, and finally, *F4* was positively correlated with *r-index* (r_S = 0.30, p<0.001); the higher *F4*, the more /r/-ful sounding the word.

B. Articulatory: gesture timing – raw and normalised lag

Fig. 4 below presents as boxplots the distribution of the *raw lag* for the significant factors: PRECEDING VOWEL (F = 42.40, p < 0.001, checked vowel estimated mean *raw lag* = -0.034s +/- 0.013s, unchecked vowel = -0.0019s +/- 0.0128s) and CLASS (F = 22.98, p < 0.001, WC estimated mean *raw lag* = 0.045s +/-0.019s, MC = -0.077s +/- 0.017s). For *normalised lag*, a significant interaction was found for PRECEDING VOWEL and CLASS (F = 11.12, p < 0.001). *t*-tests showed significance for checked and unchecked vowels in MC speech only t = 6.62, p<0.001.

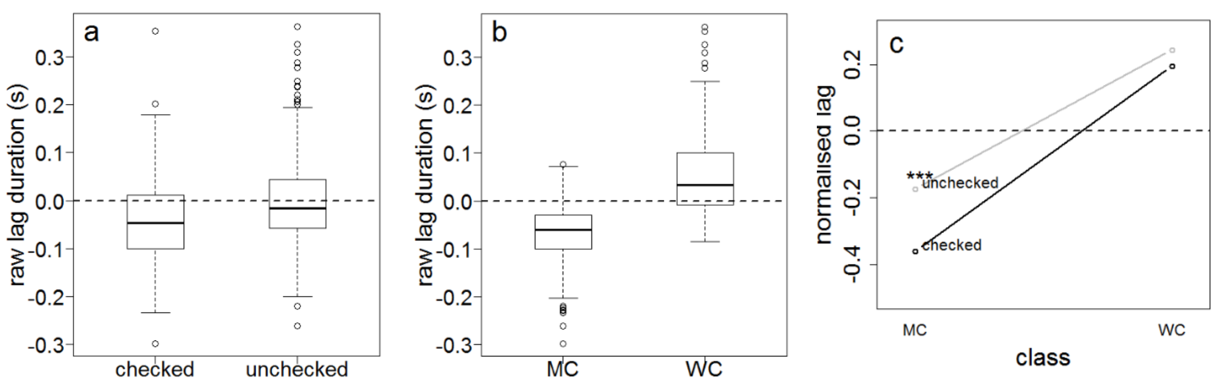

FIG. 4: Boxplots of (a) raw lag by PRECEDING VOWEL (b) raw lag by CLASS (c) interaction plot for CLASS and PRECEDING VOWEL, showing adjusted means for normalised lag. checked= /ɪ,ɛ,ʌ/, unchecked= all other vowels; MC=middle class, WC=working class. *** = significant to the <0.001 level. N=428.

In Fig. 4 above, where data points occur below the broken horizontal lines, the maximum of the /r/ gesture occurred before voicing offset in CVr words, or before the onset of the final consonant in (C)VrC words. Where data points occur above this line, some or all of the anterior lingual /r/

gesture occurred after voicing offset, or during the articulation of a following labial consonant. In the former case, we would expect /r/ tokens to be audible, in the latter case, we would expect the /r/ to be either completely inaudible, in cases where the maximum of the /r/ gesture is very delayed, or, for the /r/ to be audibly weakened to varying extents, depending on how much of the anterior lingual gesture is masked.

C. Acoustics: formant measures

1. F1

Fig. 5 below shows a significant interaction for F1 between CLASS and PRECEDING VOWEL $F = 22.11$, p = 0. MC speakers showed a lower estimated mean F1 (562Hz, +/-33Hz) than WC speakers (667Hz, +/-35Hz). /r/ after checked vowels also had a significantly higher F1 than after unchecked vowels, $t = 3.56$, $p<0.001$ for MC speakers only.

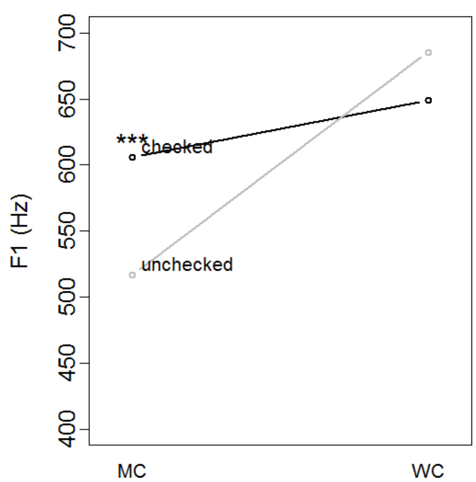

FIG. 5: Interaction plot for F1 of CLASS and PRECEDING VOWEL. N=430

1. F2

Fig. 6 shows as boxplots the distribution of F2s for the significant factors: STRUCTURE (F = 11.6, p< 0.01, CVr estimated mean F2 = 1550Hz +/- 37Hz, CVrC = 1445Hz +/- 43Hz) and SEX (F = 6.88, p< 0.05, female estimated mean F2 = 1588Hz +/- 48Hz, male = 1408 +/- 52Hz). A significant interaction for CLASS and PRECEDING VOWEL was identified (F = 9.43, p< 0.01), see Fig. 6.c, but t-tests showed no significant differences between /r/s after checked and unchecked vowels in middle-class and working-class tokens.

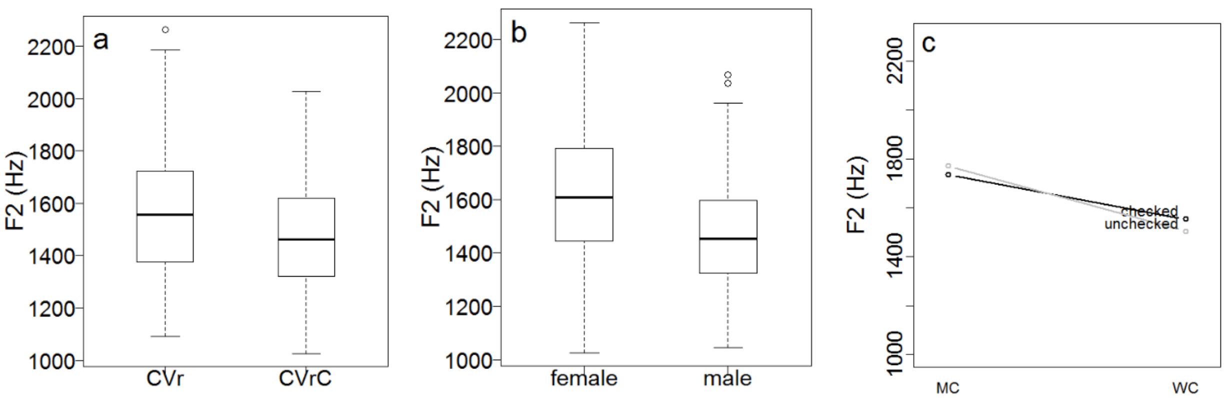

FIG. 6: Boxplots of mean F2 by (a) STRUCTURE, (b) SEX and (c) interaction plot for CLASS and PRECEDING VOWEL, showing adjusted means for F2. N=429.

2. F3

Fig. 7 below shows as boxplots the distribution of F3 for the significant factors: STRUCTURE (F = 22.75, p = 0, CVr estimated mean F3 = 2320Hz +/- 48Hz, CVrC = 2175Hz +/- 52Hz) and SEX (F = 5.27, p < 0.05, female estimated mean F3 2353 = Hz +/- 64Hz, male = 2142 +/- 69Hz). There was a significant interaction for CLASS and PRECEDING VOWEL (F = 7.61, p< 0.01), see Fig. 7.c. A t-test showed a significant difference in F3 between checked and unchecked vowels for the WC speakers only t = 3.62, p<0.001.

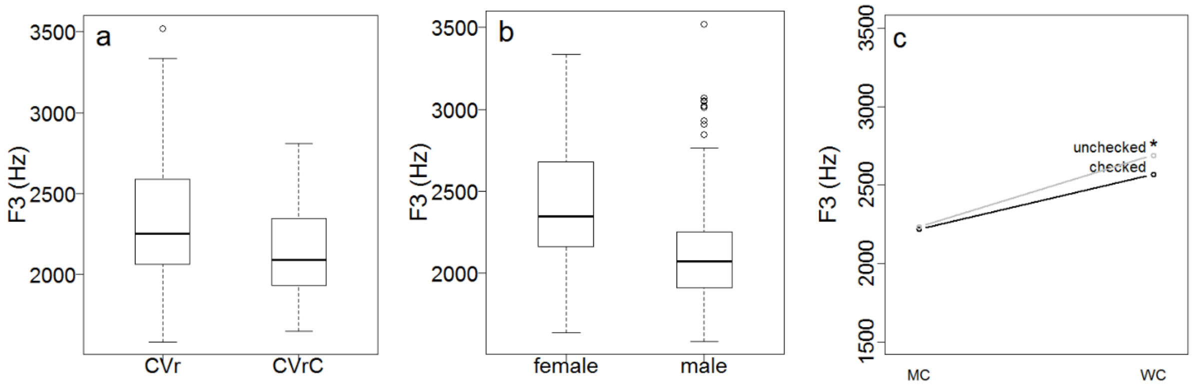

FIG. 7: Boxplots of mean F3 by (a) STRUCTURE (b) SEX, and (c) interaction plot for CLASS and PRECEDING VOWEL, showing adjusted means for F3. N=428.

3. F4

Fig. 8.a below shows as a boxplot the distribution of F4s for the significant factor: STRUCTURE (F = 4.46, p <0.05, CVr estimated mean F4 = 3845Hz +/- 63Hz, CVrC = 3757Hz +/- 70Hz). There was also a significant interaction between CLASS and PRECEDING VOWEL, (F = 4.7, p < 0.05), with a significant difference between checked and unchecked vowels in the WC group t = 3.33 , p = 0.001, Fig. 8.b.

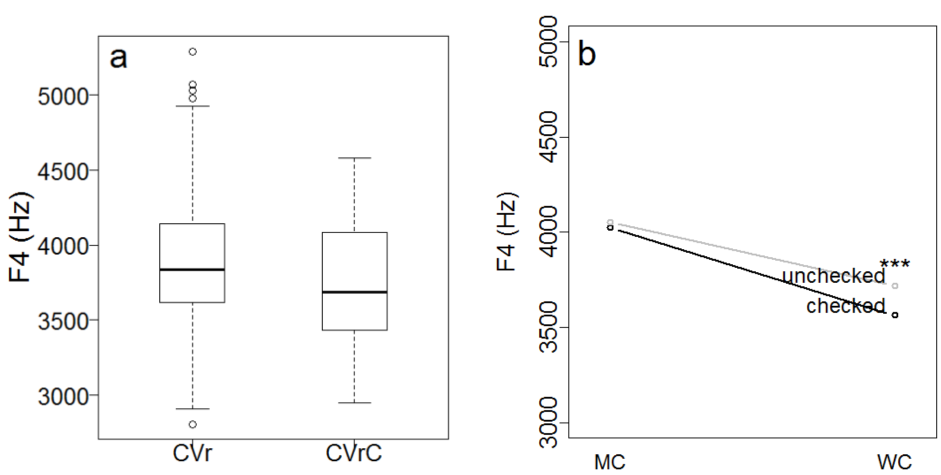

FIG. 8: Boxplots of F4 by (a) STRUCTURE, and (b) interaction plot for CLASS and PRECEDING VOWEL, showing adjusted means for F4. N=430.

D. Auditory measures: r-index score – Kruskal-Wallis

We were unable to include the ordinal *r-index* data in the mixed effects modelling; however, we carried out a series of Kruskal-Wallis nonparametric tests to explore the effect of each of our fixed factors on *r-index*. Fig. 5 below shows as boxplots the distribution of the *r-index score* for the significant factors: CLASS (χ^2 = 217.24, p<0.001), with working-class speakers heard as less /r/-ful than middle-class speakers; SEX (χ^2 = 7.68, p<0.01), with females heard as less /r/-ful than males, and PRECEDING VOWEL (χ^2 = 5.94, p<0.05), with tokens containing checked vowels heard as more /r/-ful than those containing unchecked vowels.

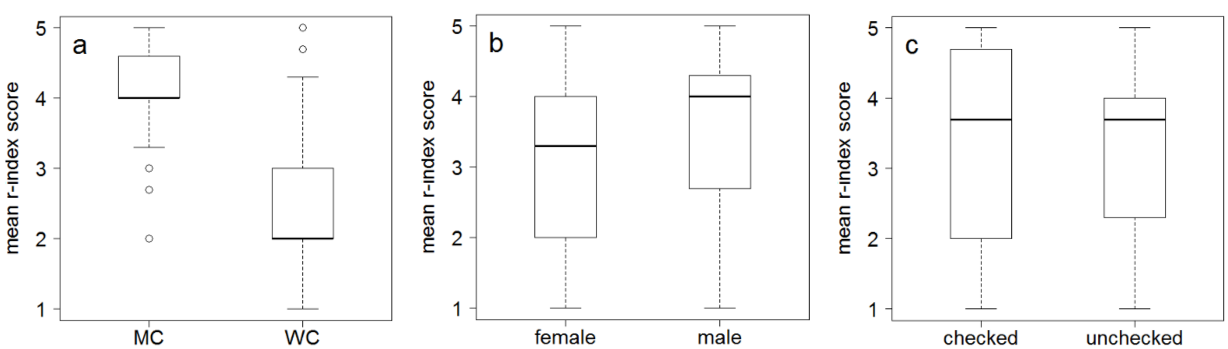

FIG. 9: Boxplots showing mean r-index scores by (a) CLASS, (b) SEX and (c) PRECEDING VOWEL. r-index category 5 = strongly /r/-ful, category 1 = no audible rhoticity. N = 408.

IV. SUMMARY AND DISCUSSION

Previous articulatory research suggested two potential mechanisms underlying the audible weakening of coda liquids: gesture reduction (Browman and Goldstein, 1995) and gesture delay (Recasens and Farnetani, 1994, Krakow, 1989, Delattre, Pierre & Freeman, Donald C., 1968).

Observations from a decade of articulatory study of postvocalic /r/ in Central Scotland, where different sociolects have markedly different strengths of rhoticity, suggested that gesture delay could be a key articulatory mechanism underlying coda /r/ weakening (Lawson et al., 2014a). Additionally, variation in gesture timing affecting coda /r/ has been observed in other languages, such as Dutch (Scobbie et al., 2009). We also know that studies of coda /l/ in English have found apical gestures for /l/ to be so delayed in some utterance-final positions that there is no evidence of them in the acoustic signal (Recasens and Farnetani, 1994). Previous articulatory studies of coda /r/ in working-class Central Scotland suggested that variation in timing might also be responsible for its weakening (Lawson et al., 2014a); however, timing variation in Scottish coda /r/ has not been quantified until now.

We hypothesised that underlying the weakly-rhotic working-class coda /r/s in our dataset, there would be a delay in the timing of the anterior lingual gesture, and we hypothesized that this variation in lingual gesture timing would correlate with variation in auditory percept of rhoticity. We also investigated the spectral characteristics of /r/ and how they correlated with gesture delay and percept of rhoticity. In the following discussion, we report on correlations between the articulatory, auditory and acoustic variables in our study and then we consider our findings in terms of our independent social and linguistic variables.

A. Correlation test findings: articulatory-auditory-acoustic relations for /r/

Correlation tests revealed the articulatory-acoustic-auditory relationship that our study aimed to uncover. Significant strong negative correlations were found between *raw*/*normalised lag* and *r-index* score, between *r-index* score and *F3*, and there was a significant strong positive correlation between *raw/normalized lag* and *F3*. There was also a strong negative correlation between *raw lag* and *F2*. In other words, the longer the *lag*, the higher *F3*, and the lower the *r-index* score (less /r/-ful sounding the word). Correlation plots also indicated that articulatory *lag*, acoustic *F2* and *F3*, and auditory *r-index* were socially stratified, with working-class speakers' data points showing longer *lags*, lower *F2*s, higher *F3*s and lower *r-index* scores than the middle-class speakers.

These findings confirm the importance of *F3*, and to a lesser extent, *F2*, in the perception of rhoticity. Tokens such as working-class speaker GWF1's "fir" [fʌˤ], shown in Fig. 2, illustrates the impact of anterior lingual gesture delay on F2 and F3 in the acoustic signal. This token was unanimously rated 2 on the rhoticity index (i.e. "derhoticised") by the authors. Close to offset of voicing, F3 is high and F2 and F3 are far apart, although there is very slight lowering of F3 and raising of F2 just before voicing offset. Ongoing formant changes (continued rising F2 and falling F3) are visible on the spectrogram thanks to a breathy exhale following voicing offset. This breathiness is not a systematic feature of working-class pronunciation, and therefore not likely to be a general means by which delayed /r/ articulations are cued, but it shows that articulatory movements associated F3 lowering and F2 raising continue after voicing has ceased. For many of the working-class tokens, only the initial stages of F3 lowering and F2 raising were present before voicing offset, or final-consonant onset. In extreme cases, all F3 lowering and F2 raising is masked by voicing offset or by a final labial consonant. This example in Fig. 2, and the findings of this study in general, helps illustrate why trained phoneticians have coded tokens of /r/ words in such different ways in auditory-only analysis of Central Scottish speech (e.g. Stuart-Smith, 2007b).

The finding that anterior lingual gesture delay results in gradient truncation of the acoustic cues associated with rhoticity might be interpreted by some in terms of a production-perception loop, whereby the speaker delays the anterior lingual gesture, part or all of the gesture becomes inaudible, the listener fails to perceive the covert /r/ and begins to produce /r/-ful words without /r/, i.e. the listener as the source of sound change (Ohala, 1981). Our research and the work of others suggests that such listener reinterpretation is not the main driver of sound change in this community. Mimicry studies carried out by the authors found that very weakly /r/-ful audio tokens were rarely mimicked as /r/-less (Lawson et al., 2011). Top-down information, such as lexical access are likely playing a part in /r/ preservation, as the only weakly /r/-ful token that was mimicked as /r/-less in (ibid.) was the only word-list stimulus that had a meaningful minimal-pair counterpart i.e. “hurt” mimicked as “hut”. However, in a subsequent mimicry experiment involving mimicry of nonsense words with weak /r/s and covert gestures, /r/-less mimicry was also rare (Lawson et al., 2014b). Lennon (2013) has shown, furthermore, that those within the Central Scotland speech community can reliably distinguish between /r/-ful and /r/-less minimal pairs, while speakers from outside of the community perform at chance level. This set of findings from mimicry and perceptual experiments perhaps go some way towards answering the perennial question of why sound change does not happen more often. It would seem that lexical access can prevent the reanalysis of /r/-ful to /r/-less when changes to gesture timing weaken the acoustic features associated with rhoticity, but also that even minimal acoustic changes in F2 and F3, or variation in prerhotic vowel duration can cue rhoticity for some listeners.

Significant correlations relating to *F4* potentially contribute to our understanding of the impact of rhotic tongue shape on acoustics. These correlations suggest that the middle-class speakers in the study, who predominantly use bunched rhotic variants, see §I.B and Fig. 1, produce rhotics with the highest F4 values. It was found that *F3* and *F4* had a significant negative correlation, that *F4* was significantly negatively correlated with lag and that *F4* was positively correlated with *r-index*. In other words, the lower *F3*, the higher *F4*, the shorter the *lag* the higher

F4, and the higher *F4,* the higher the *r-index* score (more /r/-ful the word). These correlational findings tend to support Zhou et al. (2008) regarding their conclusions concerning the acoustic characteristics of bunched /r/, namely that there is a greater distance between F3 and F4 for bunched /r/ than for retroflex /r/.

B. Social and linguistic factors in coda /r/ weakening

Speaker social class is rarely considered in phonetic, and especially lab-based, articulatory studies of consonant weakening. In fact, articulatory phonetic accounts of speech sounds are often based on analysis of "standard" or prestigious varieties of languages and on non-stratified convenience samples. Changes in the production of coda /r/ in two Central Scottish sociolects provided an ideal opportunity to assess the impact of lingual gesture timing variation in coda /r/ weakening and strengthening. Earlier articulatory studies had suggested that, in addition to social stratification of coda /r/ tongue shape, there might also be underlying variation in the timing of the anterior lingual gesture for coda /r/, with working-class speakers, in particular, delaying the anterior gesture (Lawson et al., 2014a). Due to the fact that tongue shape is socially-stratified in our dataset, see Fig. 1, we were also able to gather some evidence regarding the impact of tongue shape on formant structure of /r/.

Results confirmed our hypotheses regarding the effect of CLASS on the dependent variables studied; CLASS was found to have a significant effect for *raw* and *normalised lag* and F1 – F4. Nonparametric tests on the *r-index* dependent variable also showed that CLASS was a highly significant predictor of audible strength of /r/. Working-class speakers showed significantly greater gesture *lag*, lower *r-index* score, higher *F1*, lower *F2*, higher *F3* and lower *F4*s than middle-class speakers. The acoustic characteristics of the middle-class dataset are consistent with the findings of Heselwood (2009) regarding rhotic perceptual cues and auditory integration (Bladon,

1983) among formants 1-3. Heselwood found that perception of rhoticity depends on F2 being distant enough from F1 to avoid auditory integration, while F2 and F3 must be close enough for auditory integration to take place, resulting in a perceptual peak in the F2 region. Additionally, as mentioned in §IV.A, a study by Zhou et al. (2008) suggests that a lower F4, closer to F3, is associated with a retroflex (here 'tip/front raised'), rather than a bunched articulation for /r/. Zhou et al. obtained MRI and acoustic tokens of retroflex and bunched /r/ respectively from two adult American males. Spectral analysis showed that for a bunched /r/, there was a much greater distance between F3 and F4, and a much narrower distance between F4 and F5, compared to the retroflex acoustics. Unfortunately, variation in F5 was not reported in our study, as we were not confident of the accuracy of F5 measures, particularly in the working-class dataset where measures were often taken at voicing offset.

Social stratification of bunched and tip-up /r/ in our recordings might also explain significant interactions between the factors CLASS and PRECEDING VOWEL for normalised lag. Previous research of rhoticity in the adolescent Central Scottish speech community showed that bunched /r/ exerts a strong coarticulatory pressure on preceding checked vowels, resulting in both coalescence of the vowel and the following /r/, and the neutralisation of a three-way prerhotic contrast - /ɪr, ɛr, ʌr/ to [ɚ] (Lawson et al., 2013). This vowel+/r/ coalescence causes the point of maximum constriction of the anterior lingual gesture in /r/ to occur close to the nucleus of the syllable. We have even observed the maximum of the anterior /r/ gesture occurring as early as during aspiration in a middle-class pronunciation of the word *purr* [pʰɚ] in another Central Scottish UTI dataset. It was therefore unsurprising that the anterior lingual gesture for /r/ occurred significantly earlier in words containing checked vowels than in those containing unchecked vowels for middle-class speakers, but not for working-class speakers.

We might expect the impact of SEX, assuming smaller vocal tract cavities in girls than boys, to be reflected in an overall raising of the acoustic measures. We did find significantly higher values for females for F2 and F3, which is consistent with this assumption, but F1 and F4 showed no SEX effect. To some extent, this is likely to be due to the fact that our speakers were in early adolescence, at an age where females often temporarily outgrow their male peers, and we observed a lot of variation in height in the male and female cohorts. Listeners also often misidentified speaker gender when listening to the audio recordings. It is also possible that differences in underlying tongue shape for /r/ were having a bigger impact on vocal tract resonances than vocal tract length, see Stuart-Smith (2007) for another example of this phenomenon.

STRUCTURE was not found to affect *raw* or *normalised lag*, or *r-index.* The effect of STRUCTURE on acoustic variables was not straightforward, with the acoustic and perceptual effects of STRUCTURE appearing to partly contradict one another; significantly lower F2s were found in CVrC words, which fits in with weaker rhoticity in these words, but the significantly lower F3s in CVrC words, we would expect to be associated with stronger rhoticity. One explanation for this apparent contradiction between acoustic and auditory findings might be that, as raters heard entire word tokens, they might have had better auditory perception of delayed /r/ in CVr words resulting from formants audible after voicing offset, if noise was present, e.g. breathy exhalation, see Fig. 2, right. This kind of information would not be available for CVrC words with the complete masking effect of a final consonant.

V. CONCLUSION

This study of the phonetic basis of coda lenition identifies gesture delay as a key mechanism for coda /r/ weakening, affecting primarily the third and second formants by causing their maximum and minimum values respectively to be masked by other speech events such as voicing offset, or onset of a following consonant. Furthermore, statistical evidence concerning acoustic variables in this study, taken alongside a tongue-shape analysis of the same data (Lawson et al., 2014a), supports the findings of (Zhou et al., 2008) that underlying bunched tongue configuration is reflected in a greater separation of F3 and F4, than for tip/front-raised /r/ variants.

Our study presents a picture of /r/ weakening through change in gesture timing that causes tongue gestures in working-class speech to be partially auditorily masked, or even auditorily covert. It might be tempting to assume that further weakening will occur through a perception-production loop where perceptual reinterpretation of covert articulations occur, but mimicry studies carried out by the authors (Lawson et al., 2011b, Lawson et al., 2014b) and perceptual studies Lennon (2013) suggest that this is not happening and that speakers from these communities do not, on the whole, reinterpret weakly /r/-ful words as /r/-less. The results of linear mixed effects modelling in this study consistently showed CLASS to be the dominant predictor of variation in /r/ gesture timing and formant variation, and nonparametric tests for *r-index* show that CLASS is the most important factor for audible rhotic strength; in other words, it seems that speakers are exploiting lingual gesture timing, as well as using tongue shape, to index social information. We therefore conclude that social factors are the main driving force behind timing variation in rhotics in this speech community and we hypothesise that social factors will continue to be the driving force behind any further change, rather than the production-perception mechanism.

Acknowledgements

This work was supported by Economic and Social Research Council grants RES-062-23-3246 and ES-N008189-1.

Endnotes

[1] Qualification for free school meals is a widely-used indicator of socioeconomic deprivation, as it requires parents/guardians of students to be in receipt of income support, see (Taylor, 2017).

[2] N.B. pharyngealisation does not always have to be present for the "derhoticised" category, but often is, see (Speitel and Johnston, 1983). The "derhoticised" category refers to an audio token where there is some hint of an /r/ present, without there being a clearly audible /r/ segment.

References

Adank, P., Smits, R. and Van Hout, R. (**2004**). “A comparison of vowel normalization procedures for language variation research.” J. Acoust. Soc. Am. **116**, 3099-3207.

Aitken, A. J. (**1979**). “Scottish Speech: a historical view with special reference to the Standard English of Scotland.” In *Languages of Scotland*, edited by Aitken, A. J. and McArthur, T. (W&R Chambers Ltd., Edinburgh). Pp.85-118.

Bladon, A. (**1983**). “Two-formant models of vowel perception: shortcomings and enhancements,” Speech Communication **2**, 305-313.

Boersma, Paul & Weenink, David (**2013**). “Praat: doing phonetics by computer.” [Computer program]. Version 6.0.23, http://www.praat.org/

Browman, C. P. and Goldstein, L. (**1995**). “Gestural syllable position effects in American English,” in *Producing Speech: Contemporary Issues*, edited by Bell-Berti, F. and Raphael, L. J., (A.I.P. Press, New York). Pp. 19-33.

Core Team, R. (**2013**). “R: A language and environment for statistical computing.” [Computer program]. Version 3.3.2. https://cran.r-project.org/

Delattre, Pierre & Freeman, Donald C. (**1968**). “A dialect study of American r's by x-ray motion picture,” Linguistics **44**, 29-68.

Dobson, E. J. (**1957**). *English Pronunciation 1500 - 1700, volume II* (Oxford University Press, London). pp. 445-1073.

Drager, K. and Hay, J. (**2012**). "Exploiting random intercepts: Two case studies in sociophonetics." Lang. Var. Change **24**, 59-78.

Gamer, M., Lemon, J., Fellows, I. and Singh, P. (**2012**). "irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84."

Hay, J. (**2011**). "Statistical analysis," in *Sociophonetics: A student's guide*. Ch. 16, edited by DiPaolo, M. and Yaeger-Dror, M. (Routledge, London), pp.198-215.

Heselwood, B. (**2009**). "Rhoticity without F3: Lowpass filtering, F1-F2 relations and the perception of rhoticity in 'NORTH-FORCE', 'START' and 'NURSE' words." In Leeds Working Papers in Linguistics and Phonetics **14** (University of Leeds, Leeds), 49-64.

Jauriberry, T., Sock, R., Hamm, A. and Pukli, M. (**2012**). "Rhoticite et derhoticisation en anglais ecossais d'Ayrshire," "Rhoticity and derhoticisation in the Scottish English of Ayrshire", *Proceedings of the Joint Conference JEP-TALN-RECITAL.*, June 2012, Grenoble, France: ATALA/AFCP., pp. 89-96.

Krakow, R. A. (**1989**). "Physiological organization of syllables: a review," J. Phonetics **1**, 23-54.

Ladefoged, P. and Maddieson, I. (**1996**). "Rhotics," in *The Sounds of the World's Languages*, edited by Ladefoged, P. and Maddieson, I. (Blackwell, Oxford), pp. 215-244.

Lawson, E., Scobbie, J. M. and Stuart-Smith, J. (**2014a**). "A socio-articulatory study of Scottish rhoticity." in *Sociolinguistics in Scotland*, edited by Lawson, R. (Palgrave Macmillan, London), pp. 53-78.

Lawson, E., Stuart-Smith, J. and Scobbie, J. M. (**2014b**). "A mimicry study of adaptation towards socially-salient tongue shape variants," in University of Pennsylvania Working Papers in Linguistics **20**, Article 12.

Lawson, E., Scobbie, J. M. and Stuart-Smith, J. (**2013**). "Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation," J. Phonetics. **41**, 198-210.

Lawson, E., Scobbie, J. M. and Stuart-Smith, J. (**2011a**). "The social stratification of tongue shape for postvocalic /r/ in Scottish English," J. Sociolinguistics. **15**, 256-268.

Lawson, E., Scobbie, J.M. and Stuart-Smith, J. (**2011b**). "A single case study of articulatory adaptation during acoustic mimicry," in *Proceedings of the International Congress of Phonetic Sciences XVII,* Aug 17-21 2011, Hong Kong, pp. 1170-1173.

Lennon, R. (**2013**). "The effect of experience in cross-dialect perception: parsing /r/ in Glaswegian." Unpublished dissertation, submitted in partial fulfilment of the requirements for the degree of Master of Science in English Language and Linguistics in the School of Critical Studies, University of Glasgow.

Lennon, R., Smith, R. and Stuart-Smith, J. (**2015**). "An acoustic investigation of postvocalic /r/ variants in two sociolects of Glaswegian." *Proceedings of the international congress of phonetic sciences*, XVIII, Aug 10-15, 2015, Glasgow.

Lindau, M. (**1985**). "The story of /r/," in *Phonetic Linguistics: Essays in honour of Peter Ladefoged*, edited by Fromkin, V. (Academic Press, Orlando), pp. 157-168.

Macafee, C. (**1983**). *Glasgow. Varieties of English Around the World. Text series T3.* (Benjamins, Amsterdam). pp.1-167.

Mielke, J., Baker, A. and Archangeli, D. (**2010**). "Variability and homogeneity in American English /ɹ/ allophony and /s/ retraction." Laboratory Phonology **10**, 699-730.

Ohala, J. J. (**1981**). "The listener as a source of sound change." in *Papers from the Parasession on Language and Behaviour*, edited by Masek, C. S. and Hendrick, R. A. (Chicago Linguistic Society, Chicago), pp. 178-203.

Recasens, D. and Farnetani, E. (**1994**). "Spatiotemporal properties of different allophones of /l/: phonological implications," Phonologica 1992: Proceedings of the 7th International Phonology Meeting, 195-204.

Romaine, S. (**1978**). "Postvocalic /r/ in Scottish English: Sound change in progress," in *Sociolinguistic Patterns in British English*, edited by Trudgill, P. (Edward Arnold, London), pp. 144-158.

Scobbie, J. M. and Pouplier, M. (**2010**). "The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/," Journal of Phonetics **2**, 240-259.

Scobbie, J. M., Sebregts, K. and Stuart-Smith, J. (**2009**). "Dutch rhotic allophony, coda weakening, and the phonetics-phonology interface", in QMU Speech Science Research Centre Working Paper (QMU, Edinburgh). 3-24.

Scobbie, J. M., Wrench, A. A. and van der Linden, M. (**2008**). "Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement," in Proceedings of the 8th International seminar on speech production. Dec 8-12, Strasbourg, France, 373-376.

Scobbie, J. M., Turk, A., Hewlett, N. (**1999**). "Morphemes, phonetics and lexical items: The case of the Scottish Vowel Length Rule." *Proceedings of the XIVth International Congress of Phonetics Sciences*, Aug. 1 – 7 1999, San Francisco, U.S.A. pp. 1617-1620.

Speitel, H. H. and Johnston, P. A. (**1983**). "A Sociolinguistic Investigation of Edinburgh Speech," Unpublished ESRC final report.

Sproat, R. and Fujimura, O. (**1993**). "Allophonic variation in English /l/ and its implications for phonetic implementation," J. Phonetics **21**, 292-311.

Stuart-Smith, J. (**2007a**). "Empirical evidence for gendered speech production: /s/ in Glaswegian," in Laboratory Phonology 9, edited by Cole, J. and Hualde, J. (Mouton de Gruyter, Berlin), pp. 65-86.

Stuart-Smith, J. (**2007b**). "A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents," in *Proceedings of the International Congress of Phonetic Sciences XVI*, Aug 6 - 10 2007, Saarbrucken, Germany, p. 1307.

Stuart-Smith, J. (**2003**). "The phonology of modern urban Scots," in *The Edinburgh Companion to Scots*, edited by Corbett, J., McClure, J. D. and Stuart-Smith, J. (Edinburgh University Press, Edinburgh, U.K.), pp. 110-137.

Taylor, C. (**2017**). "The Reliability of Free School Meal Eligibility as a Measure of Socio-Economic Disadvantage: Evidence from the Millennium Cohort Study in Wales," British Journal of Educational Studies, 66, 1-23.

Turton, D. (**2014**). "Some /l/s are darker than others; Accounting for variation in English /l/ with ultrasound tongue imaging." University of Pennsylvania Working Papers in linguistics 2**,** Article 21.

Twist, A., Mielke, J. and Archangeli, D. (2007). "Are 'covert' /r/ allophones really indistinguishable?" in University of Pennsylvania Working Papers in Linguistics. Vol. 13/2. Article 16.

Wells, J. C. (**1982a**). *Accents of English 3: Beyond the British Isles.* (Cambridge University Press, Cambridge).

Wells, J. C. (**1982b**). *Accents of English 1: Introduction.* (Cambridge University Press, Cambridge).

Wrench, A. (**2012**). "Articulate Assistant Advanced User Guide," [Computer Program]. Version 2.16.12. http://www.articulateinstruments.com/

Zhang, Z., Boyce, S., Espy-Wilson, C. and Tiede, M. (**2003**). "Acoustic strategies for production of American English "retroflex" /r/", in *Proceedings of the International Congress of Phonetic Sciences XV*. Barcelona, Spain, August 3-9, 2003. Universitat Autonoma de Barcelona. pp. 1125-1128.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C. and Choe, A. (**2008**). "A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/," J. Acoust. Soc. Am. **123**, 4466-4481.